Sierra Center of Excellence Overview

DOE Centers of Excellence Performance Portability Meeting

Glendale, AZ

April 19, 2016

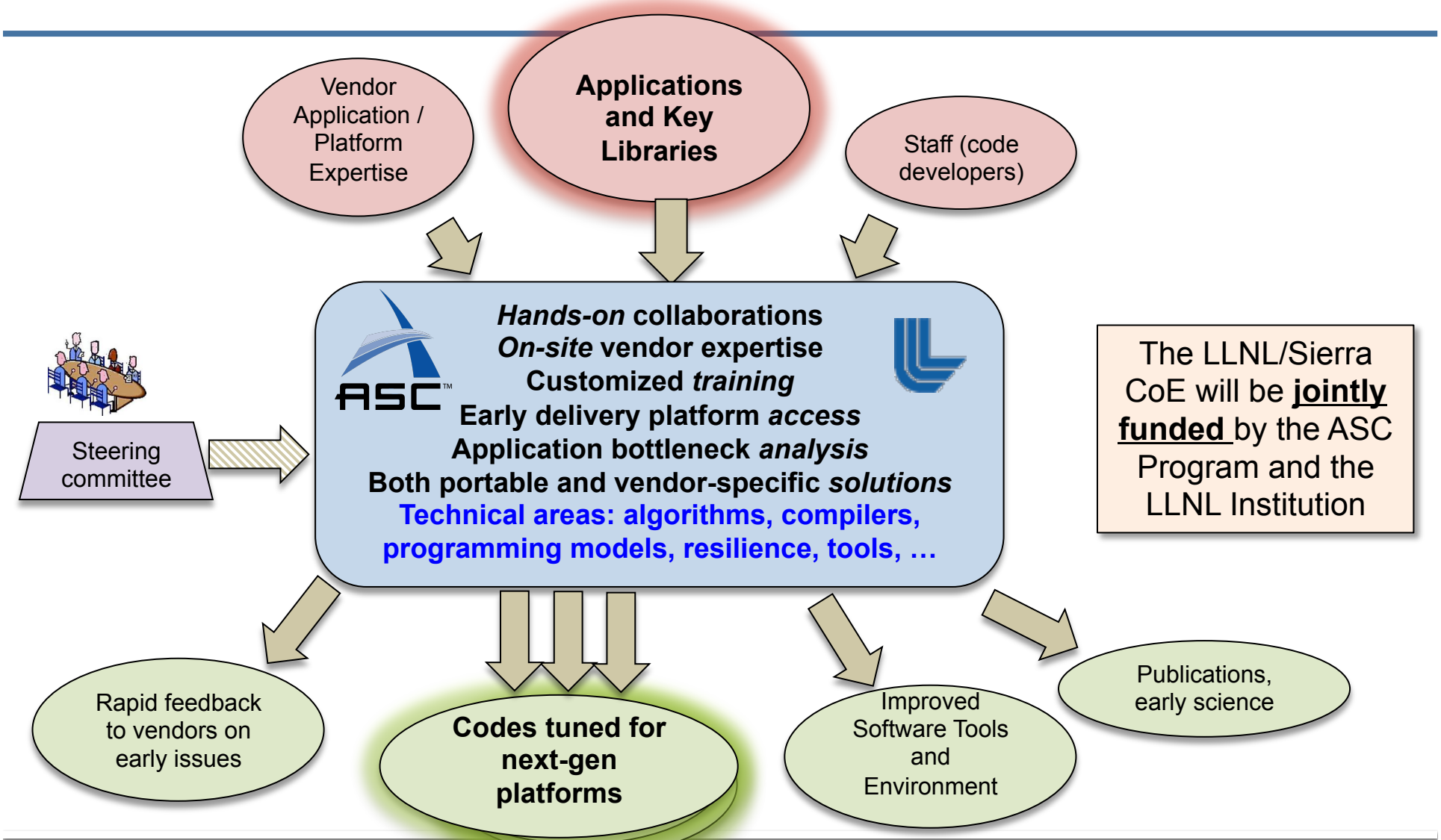
Rob Neely

LLNL-PRES-690092

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

Application enablement through the Center of Excellence – the original (and enduring) concept



ASC applications receiving COE support determined by mission-need

- ASC labs have a set of mission-critical ASC applications
  - Proxy apps (completely open, useful for initial studies)
  - Unclassified (but restricted access) applications
  - Classified applications

- ASC Apps targeting our use of Sierra were typically developed over the past 10-20 years
  - ATDM "from scratch" applications also to be included
    - Targeting first production use during lifetime of *Sierra*
- Time on the machines is granted through a tri-lab proposal process
  - CCC's – Capability Computing Campaigns
- Application preparation for new platforms is a standard part of each teams mandate
  - Multi-disciplinary teams (CS + physics + engineering expertise)
  - Progress/reviews performed through ASC milestones (L1/L2/L3)

COE vendor participation amplifies and accelerates team efforts

Starting in FY16, we expanded the *Sierra* COE to include non-ASC LLNL "institutional" applications

- LLNL has long provided institutional hardware for general unclassified use
  - Smaller versions of large classified ASC investments
  - E.g. *Vulcan* – 5PF BG/Q system (25% of *Sequoia*). Other large linux clusters
  - Planned installation of "mini-Sierra" in 2018


Vulcan (IBM BG/Q)

- Needed: similar effort to prepare institutional applications that don't receive ASC support

- ICOE (Institutional COE)
  - Mixture of internal and vendor support (code teams + AAPS team)
  - Projects selected through internal process
  - Increasing focus on data-centric computing needs
  - Managed by Bert Still

The ICOE will ensure LLNL-developed strategic codes can effectively use CORAL HW

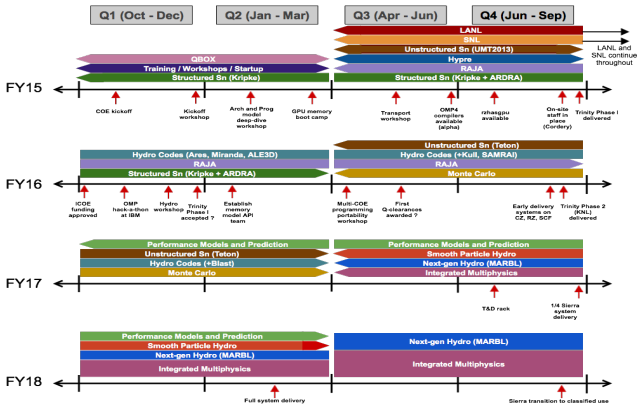
LLNL Sierra Center of Excellence Process

- Work divided into six month work plans
  - 5-6 pages of high level descriptions
  - Deliverables are optimized codes + detailed reports (available upon request)
- 2-4 full-time staff from IBM and NVIDIA
  - Primarily application-focused efforts
  - 1-2 people on site at LLNL, Q-clearances in progress
  - Remote access from IBM Research and NVIDIA
  - Ramps up over time – culminating in big push when machine arrives in 2017-18
- Institutional CoE
  - Adds an additional 2-4 vendor staff
- Coordinated with other NRE work
  - Primarily: Compiler/Tools Working Group

Details captured in a document available upon request: "*Execution Strategy for the LLNL Sierra COE*"

Our overall long-term strategy – simplified and summarized

- Codes and libraries "phased in" over time
  - Multi-year plan overlaid with hardware and compiler availability to guide work plans
- Earliest work focused on training and proxy apps
- After year one – pivoted to real applications and greater team engagement with vendor help



(No, you're not supposed to be able to read this!)

Our overarching strategy for each application (1-3 year process)

This is why this meeting is important!

Identify performance *potential* with optimized port (CUDA, alternate algorithms)

- Baseline potential performance

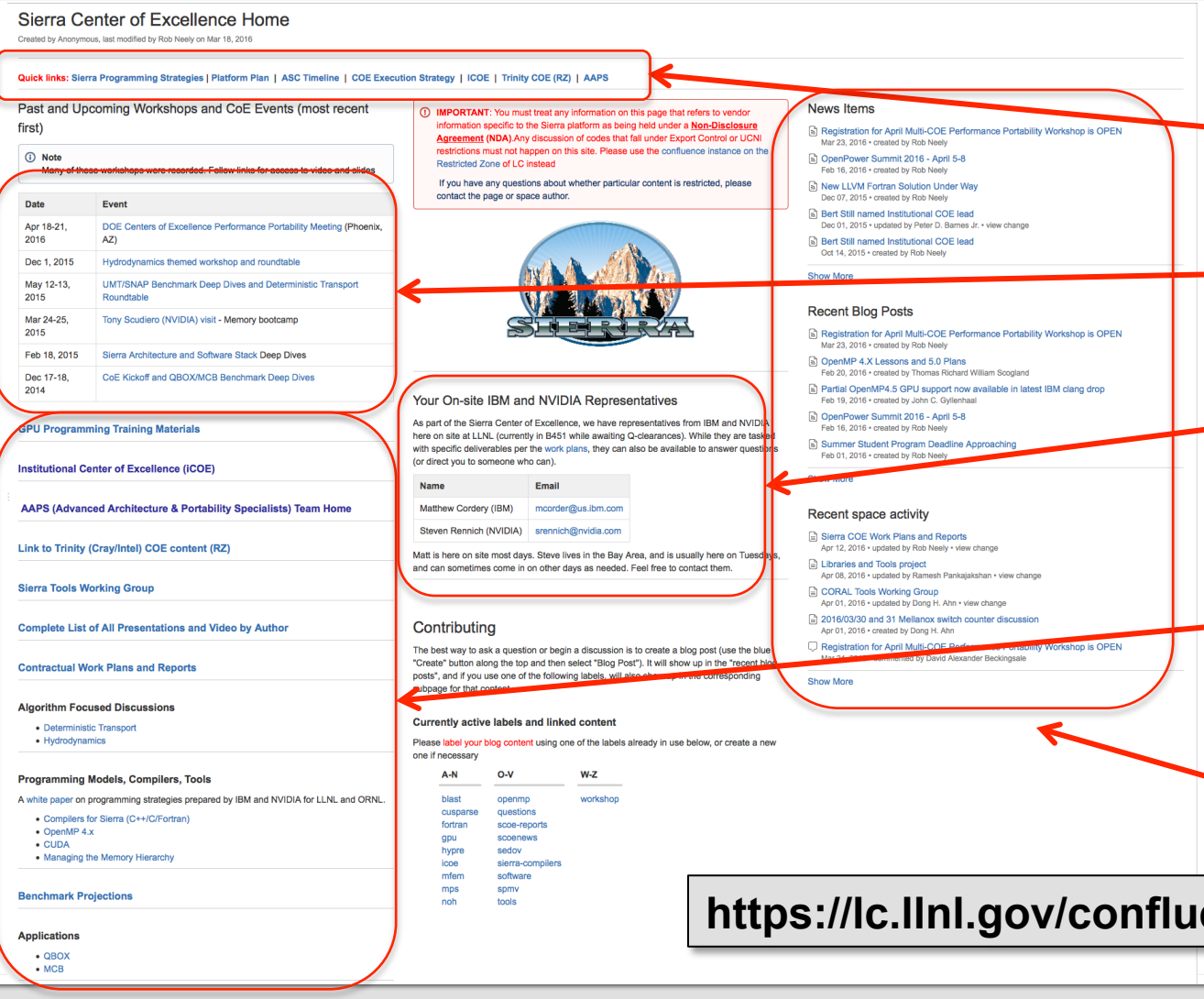
Implement in RAJA, Kokkos, OpenMP 4.x, OpenACC (Fortran)

- Compiler testing
- Establish performance impact

Ensure performance AND portability to other architectures (e.g KNL)

- Performance-portability!!!

Efforts Documented on an internal Wiki (requires LLNL authentication, NDA access)



Quick links

Workshops, training, talks, video, …

Contact info for on-site reps

Subpages by topic breakout, training, papers, CORAL benchmarks, …

Blog posts, news items

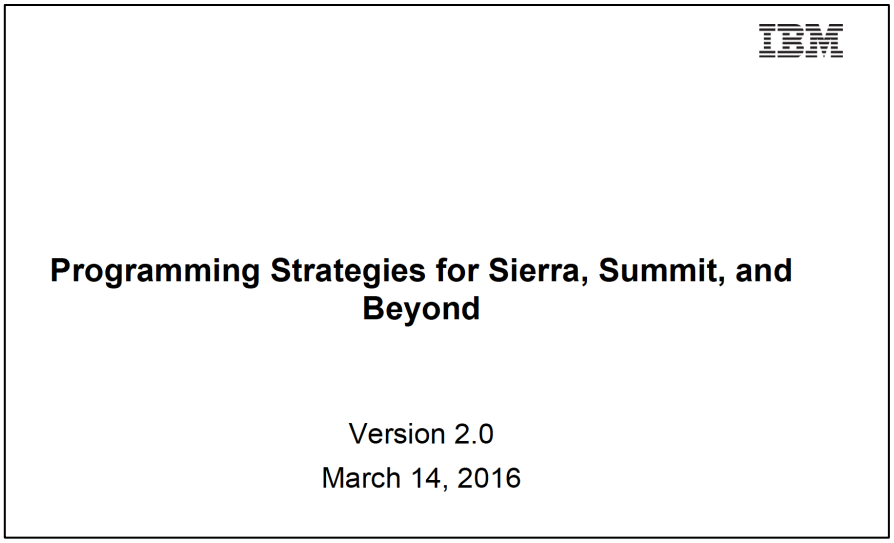
https://lc.llnl.gov/confluence/display/SCOE/

A Sampling of Sierra COE Activities

- Training (e.g. CUDA classes)
  - Accelware offered first course. Planning others (in progress)
- Talks (e.g. invited speakers) and "deep dives"
  - Many captured on video – downloadable from our website
- OpenMP4.x "hack-a-thons"
  - Hands-on experience with early compiler deliveries
  - https://codesign.llnl.gov/codesign-papers-presentations.php
- Focused workshops
  - Deterministic transport (May 2015)
  - Hydrodynamics (Dec 2015)
  - Memory hierarchy / NVLINK (planned 2016)
- Targeted application work
  - RAJA, Deterministic transport, Hydrodynamics, Monte Carlo, …

Joint white paper effort with ORNL *Summit* COE

- "*Programming Strategies for Sierra, Summit, and Beyond*"
- Excellent primer for application teams (~40 pages)
- Living document – updated every 6-12 months
- Volume I:
  - System overview
  - Programming approaches
  - Developing applications
  - Optimization
  - Examples
  - Debugging and Profiling


IBM

Programming Strategies for Sierra, Summit, and Beyond

Version 2.0

March 14, 2016

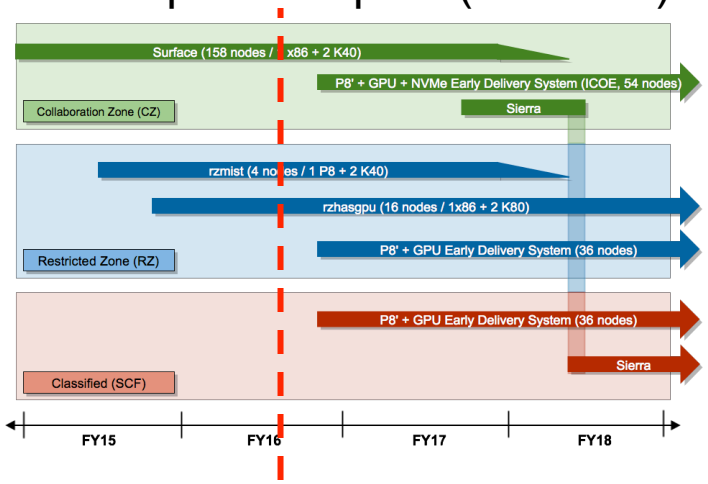
- Volume II in development
  - Will capture specific application examples

Contains **IBM/NVIDIA Confidential**
Available upon request (protect as NDA)

Summary

- 18 months under our belt
  - Halfway to delivery in FY17-18!
- Early delivery hardware anticipated this calendar year
  - 3 systems across different security zones
- LLNL teams beginning to get hands-on access to KNL hardware through Trinity COE

Sierra platform plan (tentative)



- Sierra represents the first ASC has deployed GPUs in a production system
- The COE has been critical to our efforts to get prepared
- We MUST focus on performance AND portability
- Collaborations and teaming are welcomed

Lawrence Livermore
National Laboratory